Title: The Mathematics of the Caesar Cipher

Brief Overview:

The Caesar Cipher is perhaps the most well known classical method of encrypting information. In the lessons outlined below, students will first learn how to encrypt information using a Caesar Cipher system and then examine some of the mathematical concepts which can be applied to decrypt messages that have been encrypted using a Caesar Cipher. These lessons will introduce students to the fundamentals of modular arithmetic and help students develop their problem solving skills. At the same time, these lessons will also serve to motivate a discussion of how information security has become an important part of our everyday lives (e.g., banks, Internet businesses, major corporations, and governments all protect their sensitive information using encryption).

Links to NCTM 2000 Standards:

- **Mathematics as Problem Solving**
  Students will use logic skills and the analysis of patterns and numerical data to solve enciphered messages.

- **Mathematics as Reasoning and Proof**
  Students will use inductive reasoning to devise mathematically sound approaches for encrypting and decrypting messages.

- **Mathematics as Communication**
  Students will communicate their ideas using mathematical symbols to help construct their statements and arguments about how to model the encryption/decryption process.

- **Mathematics as Connections**
  Students will discover connections among the following branches of mathematics: algebra, numerical analysis, probability, and statistics.

- **Mathematics as Representation**
  Students will learn to represent the processes of encryption and decryption using mathematical equations. They will use numerical data tables to represent the patterns within data they are examining.

- **Number and Operation**
  Students will use modular arithmetic to represent the mathematical processes they are examining.

- **Patterns, Functions, and Algebra**
  Students will use patterns and symbolic forms to develop mathematical constructs for modeling the processes of encryption and decryption.

- **Data Analysis, Statistics, and Probability**
  Students will organize and interpret numerical data gathered from the problem they are analyzing. Using this data, they will also develop and evaluate inferences, predictions, and arguments which are based on that data.

Links to Maryland High School Mathematics Core Learning Goals:

Functions and Algebra

- **1.1.1**
  Students will recognize and describe patterns to develop modular equations which model the process of encryption and decryption.

- **1.1.3**
  Students will manipulate algebraic expressions in order to evaluate modular equations.

Geometry, Measurement, and Inductive Reasoning

- **2.2.3**
  Students will use inductive and deductive reasoning to develop methods for decrypting messages.

Data Analysis and Probability

- **3.1.1**
  Students will use statistical methods to analyze frequency counts of plain text or enciphered messages.

Grade/Level:

Grades 9-12

Duration/Length:

Two 45-minute periods

Prerequisite Knowledge:

Students should have working knowledge of the following skills:

- Algebra
- Basic Arithmetic

Student Outcomes:

Students will:

- understand modular arithmetic.
- learn the basic principles of encryption.
- learn to use data analysis to help solve problems.
- use pattern recognition to express processes in a mathematical fashion.

Materials/Resources/Printed Materials:

- Pencil
- Paper

Development/Procedures:

- ### Lesson 1: Encrypting Messages Using the Caesar Cipher

A message is said to be encrypted using a substitution system if each character in the alphabet is replaced by another unique character each and every time it appears in the message. The original message is often referred to as the plain text while the encrypted message is often referred to as the cipher text.

There are many different ways to encrypt a message using a substitution system. Perhaps the most well known substitution system is the Caesar Cipher, named after the famous Roman general/statesman, Julius Caesar. Here is an example of how we would encrypt a message using a Caesar Cipher. Suppose we wanted to encrypt the following message:

```
I CAME I SAW I CONQUERED
```

We could encrypt this message by shifting every plain text character 3 places to the right to determine the corresponding cipher text character. Here is the entire table of plain text and cipher text correspondences.

```
P: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
C: D E F G H I J K L M N O P Q R S T U V W X Y Z A B C
```

Encrypting the message using the Caesar Cipher would give:

```
P: I CAME I SAW I CONQUERED
C: L FDPH L VDZ L FRQTXHUHG
```

This particular encryption scheme was actually used by Julius Caesar when he wanted to send information securely to the generals of his field armies.

Interestingly, we can represent the Caesar Cipher mathematically using modular arithmetic. Two numbers a and b are congruent modulo a number n if their difference $a - b$ is a multiple of n. Formally, we express congruence in the following way.

$$a = b\ (mod\ n)$$

Here are some examples to illustrate this new concept.

11 = 3 (mod 8) since 11 - 3 = 8 is a multiple of 8.
5 = 26 (mod 7) since 5 - 26 = -21 is a multiple of 7.

We can use modular arithmetic to study the Caesar Cipher by using numerical representations for the letters (A = 1, B = 2, ..., Z = 26). We then observe that the Caesar Cipher represents a slide of 3 letters (A goes to D, etc.), so we can represent the Caesar Cipher with the following modular equation.

$$\mathrm{C = P + 3\ (mod\ 26)}$$

In fact, one can generalize the Caesar Cipher to use a slide K, or key value, of any number (instead of 3). The generalized modular equation would then be:

$$\mathrm{C = P + K\ (mod\ 26)}$$

Historical records indicate that Julius Caesar always used a key of 3. What other possible key values could be used? (ANSWER: 0, 1, 2, ..., 25)

Hand out Problem Set 1 and allow the students to work on it in class.

- **Lesson 2: Decrypting Messages Enciphered Using a Caesar Cipher System**

We will now examine the question of how to decrypt a message encrypted with a Caesar Cipher using an unknown key value. We will learn two methods for decryption. The first is called completing the plain component. Simply put, it is exhaustion through all possible 26 key values. The second method is to use the frequency counts of the encrypted message to determine the likely key value. This second method utilizes the mathematical properties of the English language and is a much more efficient method of solution.

Method 1: Completing the Plain Component

Recall the modular equation for a Caesar Cipher system:

$$C = P + K \pmod{26}.$$

If we write this equation in terms of P, we get:

$$P = C - K \pmod{26}.$$

We will determine the proper key value by shifting the cipher text characters by all possible 26 key values. Let's consider the following example.

```
BPMZM QA UWZM BPIV WVM EIG BW AWTDM UIVG ZXWJTMUA
```

Let's take the first word of this message and write out all 26 possible shifts.
(See Handout 1.)

We see that a key value of K = 8 looks to be correct. We can verify that K = 8 is indeed the correct key by decrypting the message.

```
P: THERE IS MORE THAN ONE WAY TO SOLVE MANY PROBLEMS
C: BPMZM QA UWZM BPIV WVM EIG BW AWTDM UIVG ZXWJTMUA
```

While completing the plain component works, it can sometimes lead to a long and tedious decryption process. Let's use some new math ideas and learn how to decrypt these messages with much less effort.

Method 2: Using Frequency Counts

Suppose we are given a message. A table listing the number of times each character occurs in the message is called a table of frequency counts for that message.

Consider the following message and its associated table of frequency counts.
(See Handout 2.) What observations can be made?

Ans: Common Letters: A, E, I, N, O, R, S, T
Uncommon Letters: J, K, Q, V, W, X, Z
A, E, and I are all frequently occurring letters which are 4 letters apart.
N and O are a consecutive pair of frequently occurring letters.

R, S, and T are a consecutive triplet of frequently occurring letters.
J, K is a consecutive pair of infrequently occurring letters.
V, W is a consecutive pair of infrequently occurring letters.

We can use these observations to search for recognizable patterns of frequently (and infrequently) occurring characters in the table of frequency counts. Let's work through an example (See Handout 3.).

(Referring to Handout 3), put +'s above characters whose corresponding frequency counts are higher than expected and put -'s above characters whose corresponding frequency counts are exceptionally low (like 0 or 1). Then proceed to examine your sequence of +'s and -'s and assign the frequently and infrequently occurring letters to particular cipher text letters. This will determine the probable key value K for this message. Write out the plain text and cipher text correspondences using this key and then proceed to decrypt the message.

Hand out Problem Set 2 and allow the students to work on it in class.

Assessment:

Student progress can be easily measured by defining an appropriate scoring rubric based on the student's performance on the attached Problem Sets. One suggested rubric is listed below and is based on the point values associated to each of the two attached problem sets:

18-20 Work is complete and correct.
16-17 Work is almost complete and correct. Some minor errors may be observed.
14-15 Work is fairly complete and correct and indicates a general understanding of concepts. Noticeable errors are evident.
12-13 Some work is complete and correct. Significant errors which indicate a minimal understanding of the concepts are evident.
0-11 Work is wrong. There is no evidence that the student has grasped the concepts.

The evaluation of the problem sets should be based on:
(1) Description of mathematical concepts.
(2) Use and manipulation of mathematical equations.
(3) Use of logical reasoning.

Extension/Follow Up:

Students could work together in pairs to write, encrypt, and decrypt their own messages using the principles of the Caesar Cipher system. Particularly enthusiastic students could be encouraged to try and develop their own methods of encryption and decryption.

Author:

Paul Bellis
CMST
Howard County, MD

Problem Set 1: Encrypting Messages Using the Caesar Cipher

In problems 1-5, state whether the following are true or false.

True/False

1. 14 = 5 (mod 9) _________

2. 4 = 16 (mod 12) _________

3. 7 = 3 (mod 10) _________

4. -3 = 5 (mod 8) _________

5. 3 = 9 (mod 12) _________

6. 90 = 9 (mod 10) _________

7. Write out the correspondence between plain text and cipher text letters for a Caesar Cipher system with key value K = 5.

```
PLAIN:  A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
CIPHER:
```

8. Use the Caesar Cipher described in Problem 7 to encrypt the word ARITHMETIC.

9. Encrypt the following message using a Caesar Cipher with key value K = 18 and write a modular equation to express this system of encipherment.

MATHEMATICS IS FUN

Modular Equation:__________________________

10. Decrypt the following message, which has been encrypted using a Caesar Cipher with key value K = -1. Write out a modular equation to express this system of encipherment.

GZUD Z MHBD CZX

Modular Equation:__________________________

Solutions to Problem Set 1

1. (1 pt.) True
2. (1 pt.) True
3. (1 pt.) False
4. (1 pt.) True
5. (1 pt.) False
6. (1 pt.) False
7. (3 pts.)

```
PLAIN:  A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
CIPHER: F G H I J K L M N O P Q R S T U V W X Y Z A B C D E
```

8. (3 pts.) FWNYMRJYNH

9. (4 pts.) ESLZWESLAUK AK XMF. Modular equation: C = P + 18 (mod 26).

10. (4 pts.) HAVE A NICE DAY. Modular equation: C = P - 1 (mod 26) OR
C = P + 25 (mod 26).

Handout 1: Completing the Plain Component

SHIFT	NUMERICAL VALUES	PUTATIVE PLAIN TEXT
C	2 16 13 26 13	B P M Z M
C - 1	1 15 12 25 12	A O L Y L
C - 2	26 14 11 24 11	Z N K X K
C - 3	25 13 10 23 10	Y M J W J
C - 4	24 12 9 22 9	X L I V I
C - 5	23 11 8 21 8	W K H U H
C - 6	22 10 7 20 7	V J G T G
C - 7	21 9 6 19 6	U I F S F
C - 8	20 8 5 18 5	T H E R E
C - 9	19 7 4 17 4	S G D Q D
C - 10	18 6 3 16 3	R F C P C
C - 11	17 5 2 15 2	Q E B O B
C - 12	16 4 1 14 1	P D A N A
C - 13	15 3 26 13 26	O C Z M Z
C - 14	14 2 25 12 25	N B Y L Y
C - 15	13 1 24 11 24	M A X K X
C - 16	12 26 23 10 23	L Z W J W
C - 17	11 25 22 9 22	K Y V I V
C - 18	10 24 21 8 21	J X U H U
C - 19	9 23 20 7 20	I W T G T
C - 20	8 22 19 6 19	H V S F S
C - 21	7 21 18 5 18	G U R E R
C - 22	6 20 17 4 17	F T Q D Q
C - 23	5 19 16 3 16	E S P C P
C - 24	4 18 15 2 15	D R O B O
C - 25	3 17 14 1 14	C Q N A N
C - 26	2 16 13 26 13	B P M Z M

HANDOUT 2: Frequency Counts Example

Consider the following plain text message:

OFTEN WHEN YOU HAVE AN ENCRYPTED MESSAGE STATISTICAL
PROPERTIES CAN BE MEASURED AND UTILIZED TO HELP DECRYPT
THE MESSAGE THERE ARE MANY DIFFERENT PROPERTIES WHICH
CAN BE USED FOR THIS PURPOSE ONE SUCH STATISTIC IS THE
TABLE OF FREQUENCY COUNTS OF THE ENCRYPTED MESSAGE

Let's count the number of times that each plain text character occurs:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
15	3	11	8	38	7	3	11	12	0	0	4	5	13	11	10	1	14	19	22	8	1	2	0	6	1

Handout 3: Frequency Counts Example for Encrypted Message

Consider the following message encrypted using a Caesar Cipher with unknown key value. Examine the associated table of frequency counts and put +'s above letters with unusually high frequencies and -'s above letters with particularly low frequencies (usually 0 or 1).

P:

C: NK ZNGZ HRUCY IUGRY OT WAGXXKRY NK NGY TUZNOTM ZU

P:

C: JU COZN NGY TU XOMNZ ZU IUSVRGOT OL ZNK YVGXQY LRE

P:

C: OT NOY LGIK

Table of Frequency Counts

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	0	2	0	1	0	8	1	3	1	5	3	2	10	8	0	1	5	1	6	8	2	1	4	8	8

Using what you have learned about the patterns of frequently/infrequently occurring letters, determine the likely correspondence between plain text letters and cipher text letters. Write this correspondence below the table of frequency counts.

Use this correspondence to decrypt the message and state the correct key value.

Key Value = _____.

Answers

HE THAT BLOWS COALS IN QUARRELS HE HAS NOTHING TO DO WITH HAS NO RIGHT TO COMPLAIN IF THE SPARKS FLY IN HIS FACE.

Key Value = 6.

Problem Set 2: Decrypting Messages Using the Caesar Cipher

1. The following message has been encrypted using a Caesar Cipher with an unknown key value. Use the first word of the encrypted message to fill out the table below and complete the plain component. Then decrypt the entire message and determine the correct key value used for encryption.

DZXP XPDDLRPD NLY MP DZWGPO MJ NZXAWPETYR ESP AWLTY NZXAZYPYE

SHIFT (C - K)	NUMERICAL VALUES	PUTATIVE PLAIN TEXT
C		
C - 1		
C - 2		
C - 3		
C - 4		
C - 5		
C - 6		
C - 7		
C - 8		
C - 9		
C - 10		
C - 11		
C - 12		
C - 13		
C - 14		
C - 15		
C - 16		
C - 17		
C - 18		
C - 19		
C - 20		
C - 21		
C - 22		
C - 23		
C - 24		
C - 25		
C - 26		

Key Value K = ________.

2. The following message has been encrypted using a Caesar Cipher with an unknown key value. Use the table of frequency counts below the message to determine which cipher characters occur frequently and infrequently by placing +’s and -’s above the letters. Determine the likely key value by comparing the +’s and -’s to known patterns among frequently and infrequently occurring letters. Decrypt the entire message and state the key value used to encrypt the message.

```
P:
C: BMBLG HMTLX TLRMH WXVKR IMTFX LLTZX PAXGR HNWHG HMDGH

P:
C: PPAXK XMAXP HKWLU XZBGT GWXGW HYMXG FHKXM AHKHN ZATGT

P:
C: ERLBL BLKXJ NBKXW MHWXV KRIML NVAFX LLTZX L
```

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
6	6	0	1	1	3	9	12	2	1	8	13	10	4	0	4	0	5	0	8	1	3	7	17	1	4

Key Value K = _______.

Solutions to Problem Set 2

1. (10 pts.) SOME MESSAGES CAN BE SOLVED BY COMPLETING THE PLAIN COMPONENT

SHIFT (C - K)	NUMERICAL VALUES	PUTATIVE PLAIN TEXT
C	04 26 24 16	D Z X P
C - 1	03 25 23 15	C Y W O
C - 2	02 24 22 14	B X V N
C - 3	01 23 21 13	A W U M
C - 4	26 22 20 12	Z V T L
C - 5	25 21 19 11	Y U S K
C - 6	24 20 18 10	X T R J
C - 7	23 19 17 09	W S Q I
C - 8	22 18 16 08	V R P H
C - 9	21 17 15 07	U Q O G
C - 10	20 16 14 06	T P N F
C - 11	19 15 13 05	S O M E
C - 12	18 14 12 04	R N L D
C - 13	17 13 11 03	Q M K C
C - 14	16 12 10 02	P L J B
C - 15	15 11 09 01	O K I A
C - 16	14 10 08 26	N J H Z
C - 17	13 09 07 25	M I G Y
C - 18	12 08 06 24	L H F X
C - 19	11 07 05 23	K G E W
C - 20	10 06 04 22	J F D V
C - 21	09 05 03 21	I E C U
C - 22	08 04 02 20	H D B T
C - 23	07 03 01 19	G C A S
C - 24	06 02 26 18	F B Z R
C - 25	05 01 25 17	E A Y Q
C - 26	04 26 24 16	D Z X P

Key Value K = 11.

2. IT IS NOT AS EASY TO DECRYPT A MESSAGE WHEN YOU DO NOT KNOW WHERE THE WORDS BEGIN AND END OFTEN MORE THOROUGH ANALYSIS IS REQUIRED TO DECRYPT SUCH MESSAGES

Key Value K = 19.